

Attività A.A. 2022/2023 – Big Data

L'attività consiste nell'approfondire uno degli argomenti visti nelle lezioni relative ai modelli NOSQL, al CAP theorem e alle architetture per big data.

Esempi di possibili approfondimenti sono:

- un sistema che implementa un modello NOSQL non sperimentato a lezione
- il paradigma di computazione map-reduce
- il paradigma e le strutture dati di Spark (RDD e relative estensioni spark streaming, spark SQL , structured streaming, ecc.)
- una o più caratteristiche di un modello NOSQL non sperimentato a lezione (key-value o column based)
- confronto tra sistemi diversi che implementano lo stesso modello NOSQL
- approfondimento del CAP theorem
- supporto di uno o più sistemi NOSQL a diversi livelli di availability e/o consistency
- approfondimento di una architettura per DDS tra Hadoop e Spark
- confronto tra diverse architetture per DDS
- modalità di salvataggio delle informazioni su piattaforme BigData (ad es. parquet vs avro in Hadoop) con discussione di pro e contro e possibili casistiche di utilizzo
- librerie di colloquio verso il mondo esterno su piattaforme BigData (ad es. Sqoop, Flume, ... in Hadoop) con discussione di pro e contro e possibili casistiche di utilizzo

Possibili fonti informative, oltre a Google e alla manualistica ufficiale, sono le digital library scientifiche come google scholar <https://scholar.google.it/> e Mendeley <https://www.mendeley.com>

Al termine dell'attività di approfondimento, occorre produrre una relazione con le seguenti caratteristiche:

- la relazione deve avere un titolo e contenere un abstract (riassunto del documento), una sezione di conclusioni e un elenco di riferimenti bibliografici;
- è gradito includere una piccola sperimentazione pratica (es. un breve codice di prova per una piattaforma studiata, brevi frammenti di codice a confronto delle caratteristiche di diversi sistemi, ecc.), riassumendo nella parte finale della relazione cosa è stato provato e, se applicabile, i risultati ottenuti;
- la relazione deve essere lunga almeno 2000 parole e non più di 3000 parole.

Consegna: relazione in formato PDF ed eventuale codice prodotto da inviare via mail al docente (riccardo.martoglia@unimore.it)

Scadenza per premio partecipazione: **25 novembre 2022**